

A logistic regression model for predicting the occurrence of intense geomagnetic storms

N. Srivastava

Udaipur Solar Observatory, Physical Research Laboratory, P.O. Box 198, Udaipur, India

Received: 4 March 2005 – Revised: 16 August 2005 – Accepted: 22 August 2005 – Published: 22 November 2005

Part of Special Issue “1st European Space Weather Week (ESWW)”

Abstract. A logistic regression model is implemented for predicting the occurrence of intense/super-intense geomagnetic storms. A binary dependent variable, indicating the occurrence of intense/super-intense geomagnetic storms, is regressed against a series of independent model variables that define a number of solar and interplanetary properties of geo-effective CMEs. The model parameters (regression coefficients) are estimated from a training data set which was extracted from a dataset of 64 geo-effective CMEs observed during 1996–2002. The trained model is validated by predicting the occurrence of geomagnetic storms from a validation dataset, also extracted from the same data set of 64 geo-effective CMEs, recorded during 1996–2002, but not used for training the model. The model predicts 78% of the geomagnetic storms from the validation data set. In addition, the model predicts 85% of the geomagnetic storms from the training data set. These results indicate that logistic regression models can be effectively used for predicting the occurrence of intense geomagnetic storms from a set of solar and interplanetary factors.

Keywords. Solar physics, astrophysics, and astronomy (Flares and mass ejections) – Magnetospheric physics (Solar wind-magnetosphere interaction)

1 Introduction

The study of terrestrial consequences of earthward-directed strong CMEs is known as space weather study. It is now well established that most of the strong geomagnetic storms have their sources in fast-moving halo CMEs (Gosling et al., 1990; Srivastava and Venkatakrishnan, 2002). However, there are exceptional cases where some intense storms cannot be traced back to a single or a full halo CME, as observed by Cane and Richardson (2003); Zhang et al. (2003); Zhao and Webb (2003); Schwenn et al. (2005). Although the

space-weather prediction has been facilitated by the launch of the SoHO (Fleck et al., 1995) and furthermore by ACE (Hovestadt et al., 1995), it is still difficult to forecast the occurrence and intensity of geomagnetic storms. Also, available prediction techniques often result in false alarms, while missing some of the super-intense geomagnetic storms. Most of the available prediction schemes for the strength of the geomagnetic storms rely on inputs from the interplanetary sources of the storms and are based on the work of Burton et al. (1975). They found that the intensity of the storms largely depend on two main parameters of the solar wind, namely the solar wind speed and the southward component of the IMF. Other prediction schemes are, in fact, real-time prediction schemes (Feldstein, 1992; Lundstedt, 1992; Wu and Lundstedt 1996; Fenrich 1998; O’Brien and McPherron, 2000) which use the original formula of Burton et al. (1975). While these schemes yield generally accurate predictions, their prior warnings are only a few hours in advance of the actual occurrence of the geomagnetic storm, mainly because they rely on in-situ properties of the solar wind that can only be measured close to the Earth. For example, the ACE measurements which are made upstream at the L1 point give 30–60 min of warning time. Greater lead or warning time requires a solar wind monitor further upstream (McPherron et al., 2004). For practical applications, it is necessary to forecast space weather well in advance, so that precautionary measures can be put in place (Feynman and Gabriel, 2000). This involves prediction of a) the strength of a geomagnetic storm soon after the launch of a CME from the Sun, and b) the arrival time of the CME at the Earth. This type of advance forecasting requires the identification of key solar parameters that determine the geo-effectiveness of a CME. Amongst several characteristic features, for example, southward B_z , duration, wind speed and density, Chen et al. (1996, 1997) chose the duration and the magnitude of B_z as important quantities for predicting geo-effectiveness. However, the characteristics of the solar sources of intense/super-intense geomagnetic storms that influence their interplanetary properties and the mode of their

influence are not yet well understood. Recently, a number of studies on solar sources of severe geomagnetic storms have been carried out to understand the solar-terrestrial relationship (Feynman and Gabriel, 2000; Plunkett et al., 2001; Wang et al., 2002; Zhang et al., 2003; Vilmer et al., 2003; Schwenn et al., 2005; Srivastava, 2005). Such understanding can give us important parameters for forecasting the occurrence of intense/super-intense geomagnetic storms well in advance. Srivastava and Venkatakrishnan (2004) investigated the relationship between the solar and interplanetary parameters of the geo-effective CMEs that were responsible for producing 64 major geomagnetic storms ($D_{st} < -100$ nT) at the Earth, recorded during 1996–2002. Their investigations reveal that the intensity of geomagnetic storms is strongly related to the southward component of the interplanetary magnetic field B_z , followed by initial speeds and the ram pressure of the geo-effective CME. They obtained a high correlation coefficient (0.66) between the D_{st} index and the initial speeds of the CMEs, which indicates that the initial speed of a CME can be a useful parameter for predicting the intensity of a strong geomagnetic storm. A similarly high correlation coefficient (0.64) between the ram pressure values and the D_{st} indices indicates that the ram pressure plays an important role in the occurrence of an intense/super-intense geomagnetic storm. They also found that, although an interplanetary shock is a good predictor for the arrival of ejecta at the Earth, the shock speeds are not very reliable for predicting the resulting storm intensity. The study also suggests that the strength of a geomagnetic storm can be better predicted soon after the launch of a CME, if one can predict the ram pressure, because high ram pressure compresses the magnetic field of the magnetic cloud and intensifies the southward component of B_z , which is a good predictor of geomagnetic storms. This paper builds on the work of Srivastava and Venkatakrishnan (2004) to implement a statistical model for predicting the occurrence of intense/super-intense geomagnetic storms ($D_{st} < -100$ nT) based on the observations of 64 geo-effective events recorded during 1996–2002. Unlike previous models, which used only interplanetary measured variables for predicting the occurrence of a geomagnetic storm, the present model incorporates both solar and interplanetary variables, which were identified largely from the study of Srivastava and Venkatakrishnan (2004), to predict the occurrence of major storms.

Statistical models use large databases to establish a mathematical relation between solar and interplanetary parameters and the geomagnetic index (D_{st} index in the present study). The choice of parameters used in the prediction scheme is based on our current knowledge or understanding of physical processes. In the present study, by introducing solar parameters in the prediction scheme it is attempted to improve the “medium-term” forecasting (McPherron et al., 2004). Also, by using statistical models, one is able to ascertain the relative contribution of various parameters used in the predictive scheme.

In the following sections, we first describe, in brief, the theoretical background of the statistical model, namely a lo-

gistic regression model, used in this study. The criteria for development of a viable statistical model has been described. In subsequent sections, the actual model equation has been obtained based on both solar and interplanetary observed variables. Lastly, the testing and validation of the model has been discussed.

2 Statistical modeling for predicting geomagnetic storms

A statistical model for predicting geomagnetic storms can be defined as a highly-simplified, numeric representation of the relation between solar and interplanetary variables and the occurrence of geomagnetic storms. A generalized statistical model can be empirically represented below:

$$GMS = [f(SP_i, IP_j), P_k], \quad (1)$$

where GMS represents the occurrence of a geomagnetic storm, SP_i ($i=I$) is the i^{th} solar variable, IP_j ($j=J$) is the j^{th} interplanetary variable, P_k ($k=K$) is the k^{th} parameter of the function f that relates the solar and interplanetary variables to the occurrence of a geomagnetic storm, and I , J and K are the total numbers of solar variables, interplanetary variables and function parameters, respectively. Based on whether the relationship is hypothesized to be linear or nonlinear, a variety of linear and nonlinear functions can be used to approximate the relationship between the variables and the occurrence of a geomagnetic storm (Agresti, 1990; Finney, 1971). Linear/nonlinear dependence of the magnetospheric dynamics have been studied by several workers, for example, Johnson and Wing (2004) and references therein. While models based on machine learning (for example, artificial neural networks and Bayesian network classifiers) use nonlinear functions, several other statistical models, like linear regression and naïve Bayesian models, use linear functions. Nonlinear models generally fit the data more efficiently, but many nonlinear models (for example, neural network-based models) have a black-box type implementation, which means that the model parameters are difficult to interpret for gaining meaningful insights into the data. However, logistic regression offers a nonlinear model with the additional benefit that its parameters can be interpreted for gaining insights into the data, especially about the relative importance of various variables in predicting the occurrence of geomagnetic storms. This can help in understanding the factors that influence the geo-effectiveness of CMEs. Moreover, logistic regression is generally considered suitable for predictive modeling of dichotomous events that can be represented by a binary-state variable (Hosmer and Lemeshow, 2000). Such events generally include the presence/absence or the occurrence/non-occurrence type of events. In view of the interpretability of the model parameters and also because the objective in the present study is to predict the occurrence of an intense/super-intense geomagnetic storm, a logistic regression model for modeling the relation in Eq. (1) was selected.

Table 1. Coded values of dependent and independent variables of the logistic regression model.

Names of variables	Type of variable	Encoding		Code
		Measured Parameter	Value	
D_{st} index	Binary and dependent	D_{st}	–200 to –100 nT <–200 nT	0 1
Halos	Binary and independent	Full halos Partial None	360° angular span >140° angular span	1 0 –1
Location	Binary and independent	Location-bin	Within ±40° latitude ±40° longitude Outside ±40° latitude ±40° longitude	1 0
Association with other activity	Binary and independent	Flare-bin	Flares EPs	1 0
Initial Speeds	Numeric and independent	V_i	Value in km s ^{–1}	–
Southward IMF	Numeric and independent	B_z	Value in nT	–
Total IMF	Numeric and independent	B_T	Value in nT	–
Ram pressure	Numeric and independent	P_R	Value in dynes cm ^{–2}	–

2.1 Logistic regression

The objective of modeling the relation between the occurrence of an intense/super-intense geomagnetic storm (dependent variable) and solar and interplanetary variables (independent variables) is to estimate the probability of the former, given the incidences of the latter. Since the probability of an event must lie between 0 and 1, it is impractical to model probabilities with linear regression techniques, because the linear regression model allows the dependent variable to take values greater than 1 or less than 0. The logistic regression model is a type of generalized linear model that extends the linear regression model by linking the range of real numbers to the 0–1 range. Consider the existence of an unobserved continuous variable, Z , which can be thought of as the “propensity towards” the occurrence of an intense/super-intense geomagnetic storm, with larger values of Z corresponding to greater probabilities of the occurrence of an intense/super-intense geomagnetic storm. Mathematically,

$$\Pi_i = \frac{e^{Z_i}}{1 + e^{Z_i}} = \frac{1}{1 + e^{-Z_i}} \quad (2)$$

$$Z_i = \log\left(\frac{\Pi_i}{1 - \Pi_i}\right), \quad (3)$$

where Π_i is the probability of the occurrence of intense/super-intense geomagnetic, given the i^{th} observation of the solar and interplanetary variables, and Z_i is the value of the continuous variable Z . In the logistic regression model, the relationship between Z and the probability of the event of interest is described by the above link function (Eq. (2)). If Z were observable, a simple linear regression can be fitted to Z . However, since Z is unobserved, the predictor (or independent) variables have to be related to the

probability of the occurrence of an intense/super-intense geomagnetic storms by substituting for Z as follows (assuming that Z is linearly related to the predictors):

$$Z_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij} \quad (4)$$

$$\Pi_i = \frac{1}{1 + e^{-Z_i}}, \quad (5)$$

where b_j ($j=0$ to J) are the model parameters (known as regression coefficients), x_{ij} ($i=0$ to I ; $j=0$ to J) are the independent variables, I and J are the total number of observations. The regression coefficients are estimated through an iterative maximum likelihood method. In the above equations, Z is estimated as a natural logarithm of the odds of the occurrence of an intense/super-intense geomagnetic storm and, therefore, logistic regression estimates the probability or success over the probability of failure.

2.2 Datasets used

For implementation of logistic regression, we used the same dataset that was used by Srivastava and Venkatakrishnan (2004). It comprises solar sources and interplanetary data regarding 64 geo-effective CMEs which gave rise to intense to super-intense geomagnetic storms during 1996–2002. The dataset was created from observations of SoHO and ACE, in particular from instruments like EIT, LASCO and CELIAS aboard SoHO (Brueckner et al., 1995; Delaboudiniere et al., 1995; Hovestadt et al., 1995), and also observations from SWEPAM and MAG instruments data aboard the ACE spacecraft (Stone et al., 1998). The details can be found in Srivastava and Venkatakrishnan (2004). Out of the 64 geo-effective events, 9 were not included in the regression analysis because the data values of some of the variables for these

Table 2. Estimates of the parameters of the model (maximum likelihood.)

Variable	Estimates	Std. dev.	χ^2	Pr. $> \chi^2$
Intercept	-4.575	1.871	5.977	0.014
Halo-Bin	0.489	0.786	0.387	0.534
Flare-bin	0.506	1.008	0.252	0.615
Location-bin	0.305	1.170	0.068	0.795
V_i	0.001	0.001	0.922	0.337
B_T	-0.102	0.092	1.213	0.271
B_z	-0.243	0.109	4.937	0.026
P_R	2394 103.2	5 614 666	0.182	0.670

events were missing. The remaining 55 events were divided into two sets, namely, a training set and a validation set. The training dataset, comprising of 46 events, which included 16 super-intense storms and 30 intense geomagnetic storms, was used to train a logistic regression model. The trained model was validated on the remaining 9 events, which included 4 super-intense and 5 intense geomagnetic storms.

2.3 Model variables

The intensity of the geomagnetic storm associated with each of the 64 CMEs in the dataset is represented by the D_{st} index, which is used as the dependent variable in logistic regression. However, it was converted into a binary variable, using a D_{st} index value of -200 nT as the threshold. The geomagnetic storms with $D_{st} < -200$ nT were considered super-intense and coded as 1, while the storms with -200 nT $< D_{st} < -100$ nT were considered intense and coded as 0 (Table 1). The solar properties of the geo-effective CMEs in the dataset are defined by the following variables: 1) latitude of the origin of the CME, 2) longitude of the origin of the CME, 3) flare/prominence association, 4) association with full/partial/no halo CME, and 5) initial speeds of the CME (V_i). In order to reduce redundancy, the two locational variables, namely, latitude and longitude, were combined to generate a single binary variable termed source location. Similarly, two other solar variables, viz., flare/prominence association and association with full/partial/no halo CME, were also converted into binary variables. Our study (Srivastava and Venkatakrishnan, 2004) has shown that a large number of the intense storms are associated more with flares than with prominences (75% to 25%). This implies that flares do play an important role as source regions of intense geomagnetic storms. This is also supported by the result that flares generally occur in active regions and that the magnetic energy of the source active region dictates, to some extent, the speed of the ensuing halo CMEs (Venkatakrishnan and Ravindra, 2003). Thus, the knowledge of the association of the CME with a flare or eruptive prominence indirectly suggests the magnetic energy involved, as well as the speeds of the CMEs.

The rules used for generating the binary variables are given in Table 1. The fourth solar variable, viz., initial speeds of the CME (V_i), was used as such. In all, four solar variables were used as independent variables in the logistic regression.

The interplanetary properties of the geo-effective CMEs in the dataset are described by the following measured and derived variables: 1) shock speeds (V_{SH}), 2) ram pressure (P_R), 3) total value of the IMF (B_T), 4) southward component of the IMF (B_z), 5) solar wind speeds before and after the shock (V_1 and V_2), 6) densities before and after the shock (n_1 and n_2), and 7) solar wind-magnetospheric coupling parameter ($V B_z$). Out of the above interplanetary variables, only ram pressure (P_R), total value of the IMF (B_T) and southward component of the IMF (B_z) were used as independent variables in the logistic regression. Shock speed (V_{SH}) was not used because it has a poor Pearson's correlation coefficient (0.28) with the D_{st} index (Srivastava and Venkatakrishnan, 2004). Shock speed (V_{SH}) is dependent on the solar wind speeds (V_1 and V_2) and also densities n_1 and n_2 before and after the shock and therefore much of the information contained in the variables V_1 , V_2 , n_1 and n_2 may be considered redundant. Similarly, ram pressure, which is dependent on the solar wind speed and density after the shock and shows a good correlation coefficient with the D_{st} index, was included in the regression analysis. The solar wind-magnetospheric coupling parameter ($V B_z$) is a function of the southward component of the IMF (B_z), and therefore was considered redundant for regression analysis.

2.4 Model training

The logistic regression was trained on the training dataset using XLSTAT software (<http://www.xlstat.com>). Training comprised estimation of regression coefficient using an iterative maximum likelihood method. The regression equation obtained after training the model is given below:

$$P = \frac{1}{(1 + \exp(-Z))}, \quad (6)$$

where

$$Z = (-4.57 + 0.488 \times \text{Halo-bin} + 0.51 \times \text{Flare-bin} + 0.30 \times \text{Location-bin} + 7.44 \times E-04 \times V_i - 0.24 \times B_z - 0.10 \times B_T + 2394103.2 \times P_R). \quad (7)$$

The details of the estimates of regression coefficients and a number of statistical properties of the model variables are given in Table 2.

2.5 Model validation

The logistic regression model was validated by using the regression equation (Eq. (4)) to predict the occurrence of intense/super-intense geomagnetic storms for the validation dataset. A threshold value of 0.500 was used to make a classification: if the predicted value is more than 0.500, the event

Table 3. Validation of the logistic regression model.

Data-sets		Observed	Predicted	% Correct prediction
Training	Super-Intense	16	10	62.5%
	Intense	30	29	97%
	Total	46	39	85%
Validation	Super-intense	4	2	50%
	Intense	5	5	100%
	Total	9	7	77.7%

is classified as a super-intense geomagnetic storm (coded as 1), otherwise as an intense geomagnetic storm (coded as 0).

3 Results and discussion

The results of the classification of the training and validation events using the trained logistic regression model are given in Table 3. The results show that the logistic regression model correctly classifies 62.5% of the training super-intense geomagnetic storms and 97% of the training intense geomagnetic storms. Amongst the validation events, the model correctly classifies 50% of the super-intense geomagnetic storms and 100% of the intense geomagnetic storms. The above results indicate that the model is only moderately successful in predicting super-intense geomagnetic storms. This indicates that the geo-effective CMEs that give rise to super-intense geomagnetic storms, have certain distinctive solar and/or interplanetary characteristics that have not been properly understood. As discussed in Sect. 1, the characteristics of the solar sources of intense/super-intense geomagnetic storms that influence their interplanetary properties and the mode of their influence are not yet well identified. The solar parameters used in the present model have been identified on the basis of current understanding, which may not adequately explain the observed data. More work needs to be done in order to augment or refine the independent model variables, especially the solar parameters, for a more accurate prediction of super-intense storms. Nevertheless, prediction rates of 50 to 60% for super-intense geomagnetic storms indicate the efficiency of the model in predicting space weather. The Chi-squared values of the estimates of the regression coefficients for different variables indicate that B_z is the most significant variable for predicting the occurrence of intense and super-intense geomagnetic storms, followed by B_T and V_i . The solar variables, in general, contribute relatively less in the prediction. Amongst the interplanetary variables, B_z is the most important predictor, followed by B_T and ram pressure. Amongst the solar variables, V_i is the most important predictor followed by full halo association, and association of flare/prominence.

4 Conclusions

A simple logistic regression model was implemented for predicting the occurrence of intense/super-intense geomagnetic storms based on a number of solar and interplanetary variables. The results indicate that the model can be used for predicting the occurrence of intense geomagnetic storms, although it is only moderately successful in predicting super-intense storms. However, to really validate the capability of the model, the validation data set should also use periods without any storms, for which more work is needed. The results also indicate that interplanetary variables are better predictors of the occurrence of geomagnetic storms, while the solar variables, in general, contribute relatively less in the prediction.

Future Research: The estimated logistic regression model does not predict the magnitude of the storm or the D_{st} index. Future research will be directed at estimating the strength of the storms by using more and accurately measured quantities and also by including the data for events where the strength of the storms was lower than reported here. It is also aimed to increase the number of geo-effective events by extending the dataset up to 2004. This would lead to better forecasting of space weather prediction.

Acknowledgements. We acknowledge the LASCO/EIT and CELIAS operational teams on SoHO and ACE for the data used in this paper. The author also acknowledges the support of European COST724 action for providing partial financial support to present this work at the first European Space Weather week held during November 2004 at ESTEC, The Netherlands.

Topical Editor T. Pulkkinen thanks D. F. Webb and another referee for their help in evaluating this paper.

References

- Agresti, A.: Categorical data analysis, John Wiley and Sons, New York, 1990.
- Brueckner, G. E., Howard, R. A., and Koomen, M. J., et al.: The Large Angle Spectroscopic Coronagraph, (LASCO), *Solar Phys.*, 162, 357–402, 1995.
- Burton, R. K., McPherron, R. L., and Russell, C. T.: An empirical relationship between interplanetary conditions and D_{st} , *J. Geophys. Res.*, 80, 14 204–14 214, 1975.
- Cane, H. V. and Richardson, I. G.: Interplanetary coronal mass ejections in the near-Earth solar wind during 1996–2002, *J. Geophys. Res.*, 108, 1156, doi:10.1029/2002JA009817, 2003.
- Chen, J., Cargill, P. J., and Palmadesso, P. J.: Real-time identification and prediction of geoeffective solar wind structures, *Geophys. Res. Lett.*, 23, 625–628, 1996.
- Chen, J., Cargill, P. J., and Palmadesso, P. J.: Predicting solar wind structures and their geoeffectiveness, *J. Geophys. Res.*, 102(A7), 14 701–14 720, 1997.
- Delaboudiniere, J.-P., Artzner, G. E., and Brunaud, J., et al.: Extreme Ultraviolet Imaging telescope for the SoHO mission, *Solar Phys.*, 162, 291–312, 1995.
- Feldstein, Y. I.: Modelling of the magnetic field, *Space Sci. Rev.*, 59, 83–165, 1992.

- Fenrich, F. R. and Luhmann, J. G.: Geomagnetic response to magnetic clouds of different polarity, *Geophys. Res. Lett.*, 25, 2999–3002, 1998.
- Feynman, J. and Gabriel, S. B.: On space weather consequences and predictions, *J. Geophys. Res.*, 105, 10 543–10 564, 2000.
- Finney, D. J. : *Probit Analysis*. 3rd Ed., Cambridge, London and New-York, 1971.
- Fleck, B., Domingo, V., and Poland, A.: The SoHO Mission, *Solar Phys.*, 162, 1–2, 1995.
- Gosling, J. T., Bame, S. J., McComas, D. J., and Phillips, J. L.: Coronal mass ejections and large geomagnetic storms, *Geophys. Res. Lett.*, 17, 901–904, 1990.
- Hosmer, D. W., Jr. and Lemeshow, S.: *Applied Logistic Regression*, Wiley, 2000, ISBN: 0-471-35632-8.
- Hovestadt, D., Hilchenbach, M., and Bürgi, A., et al.: CELIAS Charge Element and Isotope analysis system for SoHO, *Solar Phys.*, 162, 441–481, 1995.
- Johnson, R. and Wing, S.: A solar cycle dependence of non-linearity in magnetospheric activity, *J. Geophys. Res.*, 110, doi:10.1029/2004JA010638, 2004.
- Lundstedt, H.: Neural networks and predictions of solar-terrestrial effects, *Planet. Space. Sci.*, 40, 457–464, 1992.
- McPherron, R. L., Siscoe, G., and Arge, N. : Probabilistic Forecasting of the 3h ap Index, *IEEE transactions on Plasma Science*, 32, 1425–1438, 2004.
- O'Brien, T. Paul and McPherron, R. L.: An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay, *J. Geophys. Res.*, 105, 7707–7720, 2000.
- Plunkett, S. P., Thompson, B. J., St. Cyr. O. C., and Howard, R. A.: Source regions of coronal mass ejections and their geomagnetic effects, *J. Atmos. Sol. Terr. Phys.*, 63, 389–402, 2001.
- Schwenn, R., Dal Lago, A., Huttunen, E., and Gonzalez, W. D.: The association of coronal mass ejections with their effects near the earth, *Ann. Geophys.*, 23, 1033–1059, 2005, **SRef-ID: 1432-0576/ag/2005-23-1033**.
- Srivastava, N. and Venkatakrishnan, P.: Relation between CME speed and geomagnetic storm intensity, *Geophys. Res. Lett.*, 29, 1287–1291, 2002.
- Srivastava, N. and Venkatakrishnan, P.: Solar and interplanetary sources of major geomagnetic storms during 1996–2002, *J. Geophys. Res.*, 109(A010103), doi:10.1029/2003JA010175, 2004.
- Srivastava, N.: Predicting the occurrence of super-storms, *Ann. Geophys.*, 23, 2989–2995, 2005.
- Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., and Snow, F.: The Advanced Composition Explorer, *Space Sci. Rev.*, 86, 1/4, 357–408, 1998.
- Venkatakrishnan, P. and Ravindra, B.: Relationship between CME speeds and magnetic energy of active regions, *Geophys. Res. Lett.*, 30, 2181–2184, 2003.
- Vilmer, N., Pick, M., Schwenn, R., Ballatore, P., and Villain, J. P.: On the solar origin of interplanetary disturbances observed in the vicinity of the earth., *Ann. Geophys.*, 21, 847–862, 2003, **SRef-ID: 1432-0576/ag/2003-21-847**.
- Wang, Y. M., Ye, P. Z., Wang, S., Zhou, G. P., and Wang, J. X.: A statistical study on the geo-effectiveness of earth-directed coronal mass ejections from March 1997 to December 2000, *J. Geophys. Res.*, 107, 1340, doi:10.1029/2002JA009244, 2002.
- Wu, J.-G. and Lundstedt, H.: Prediction of geomagnetic storms from solar wind data using Elman recurrent neural networks, *Geophys. Res. Lett.*, 23, 319–322, 1996.
- Zhang, J., Dere, K. P., Howard, R. A., and Bothmer, V.: Identification of solar sources of major geomagnetic storms between 1996 and 2000, *Astrophys. J.*, 582, 520–533, 2003.
- Zhao, X. P. and Webb, D. F.: Source regions and storm effectiveness of frontside full halo coronal mass ejections, *J. Geophys. Res.*, 108, 1234, doi:10.1029/2002JA009606, 2003.